KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Recognizing single-peaked preferences on aggregated choice data

Smeulders B.

KBI_1427

Recognizing Single-Peaked Preferences on Aggregated Choice Data

Smeulders, B.*

Abstract

Single-Peaked preferences play an important role in the social choice literature. In this paper, we provide necessary and sufficient conditions for observed behaviour to be consistent with a mixture model of single-peaked preferences for a given ordering of the alternatives. These conditions can be tested in time polynomial in the number of choice alternatives. In addition, algorithms are provided which identify the underlying ordering of choice alternatives if ordering is unknown. These algorithms also run in polynomial time, providing an efficient test for the mixture model of single-peaked preferences.

1 Introduction

Preferences play an important role in many areas of research. When faced with different alternatives, be it different cars, candidates in an election, budgets, etc., it is commonly assumed that people have a preference ordering over all of these alternatives, ranking them from best to worst. Often the nature of the alternatives restricts the possible preferences in some sense. An important such restriction is given by *single-peakedness*, introduced by Black [3]. Suppose a linear ordering exists, which ranks all alternatives along a line. An agent's preferences are then single-peaked if he has a most preferred alternative, the *peak*, and when comparing two alternatives that are both on the same side of the peak, the alternative closest to the peak is preferred. This restriction is very natural when considering a situation where a single attribute of the alternatives drives the choice, for example, an election where candidates range from left to right wing or choices over budgets of various sizes. Given these examples, it is no wonder that this restriction has gained central importance in the areas of political science and social choice. Apart from being an appealing model in these areas, the assumption of single-peaked preferences has led to interesting theoretical results. For example, aggregation of single-peaked preferences avoids the Condorcet-paradox.

Given the importance of single-peaked preferences, it is of interest to test if and in which situations agents hold such preferences. Given the complete preference profile of agents Bartholdi and Trick [2] provide a polynomial time algorithm to test whether these are single-peaked in regards to some ordering of the alternatives and to identify this ordering. Escoffier et al. [5] provide a different algorithm for the same problem with a better worst-case bound. Ballester and Haeringer [1] give two forbidden substructures, whose absence is a necessary and sufficient condition for the given preference profile to be consistent with single-peakedness. Furthermore, Trick [11] provides an algorithm for recognizing single-peakedness on trees, which again runs in polynomial time. Finally, Knoblauch [7] investigates a closely related preference restriction, one-dimensional Euclidean preference profiles, and also provides a polynomial time algorithm.

In this paper, our goal is also to provide the means to test whether reported preferences are consistent with single-peaked preferences, but instead of using the full preference profile of

*ORSTAT, Faculty of Economics and Business, KU Leuven.
This research has been partially funded by FWO grant G.0447.10, and by the Inter-university Attraction Poles Programme initiated by the Belgian Science Policy Office

the agents, we will be working with aggregated preference data. For each pair of alternatives, the proportion of the population preferring one over the other is known, but there is no further information on the individual agents. The main question is then whether it is possible that a population of agents holding single-peaked preferences have these aggregated preferences. Our motivation for using this type of data is twofold. First, this setting is mathematically equivalent to one often used in choice behaviour research. In this field, it is often observed that people make inconsistent choices. When faced with repetitions of the same binary choice, choice reversals are commonplace, even within short timeframes. One possible explanation is that while at any given point in time, persons have consistent preferences, these preferences may change often. It is thus hypothesized that a person has a set of preferences, and a probability distribution over them, which gives the probability that he makes a decision based on a particular preference. Within this literature, this is known as a *mixture model* [9]. In the case of general linear preferences, testing this mixture model corresponds to testing whether the observed data lies within the linear ordering polytope [10]. Necessary and sufficient conditions, testable in polynomial time, are given for a mixture model of heuristic choice behaviour by Davis-Stober [4]. We wish to provide a test of the mixture model when the preferences are restricted to single-peaked preferences. Second, when returning to a social choice setting, the advantage of this mixture model setting is that aggregated choice data is easier to obtain than full preference profiles. Furthermore, it may not even be the case that a fixed preference order for agents is a reliable representation of the agent's preferences. One drawback of the mixture model is that it is less restrictive than tests on full preference profiles, since it is possible that the aggregated preference of a non-single-peaked population is the same as one from a single-peaked population.

The main contributions of this paper are as follows.

- Given an ordering of the alternatives, we provide necessary and sufficient conditions for testing whether aggregated preferences are consistent with a mixture model of single-peaked preferences. These conditions can be tested in polynomial time.
- We provide an algorithm which given the aggregated preferences, provides an ordering of the alternatives for which the mixture model is satisfied, if such an ordering exists.

The rest of this paper is organized as follows. In section 2, we further define single-peaked preferences and the mixture model. Section 3 contains our main result, necessary and sufficient conditions for a mixture model of single-peaked preferences to hold. Next, section 4 provides further results, specifically two algorithms to identify the underlying ordering of the alternatives. Finally, section 5 concludes.

2 Preliminaries and Notation

Consider a set A, consisting of n alternatives, and a dataset $P = \{p_{ij} \geq 0, \forall i, j \in A\}$. The values p_{ij} represent the probability that i is chosen over j. As we assume strict preferences, $p_{ij} + p_{ji} = 1$. There also exists an ordering of alternatives in A along an axis. This ordering is complete and transitive and is denoted by $\gg$. Preference orderings over the alternatives are represented by the relation $\succ$. These relations are also complete and transitive. We will use the index m to denote a particular preference ordering.

Definition 1. *A preference ordering m is single-peaked with respect to a given ordering of the* *alternatives $\gg$ if and only if for every triple $i, j, k \in A$ we have:*

$$\textit{if } i \gg j \gg k \textit{ and } i \succ_m j \textit{ then } i \succ_m k \tag{1}$$
$$\textit{if } i \gg j \gg k \textit{ and } k \succ_m j \textit{ then } k \succ_m i \tag{2}$$

The set of all preference orderings that are single-peaked in regards to an ordering $\gg$ is $O^{\gg}$. We further consider the subsets $O^{\gg}_{ij}$, for which $m \in O^{\gg}_{ij}$ if $m \in O^{\gg}$ and $i \succ_m j$. A mixture model of preference assumes that when a decision maker is faced with a choice, each preference ordering has a certain probability of being used to make the choice. When these probabilities are consistent with the numbers p_{ij}, we say that the model rationalizes the observed data.

Definition 2. *A dataset P can be rationalized by a mixture model of single-peaked linear ordering preferences with respect to a given ordering of alternatives $\gg$ if and only if there exist numbers $x_m \geq 0, \forall m \in O^{\gg}$ for which*

$$\sum_{m \in O^{\gg}_{ij}} x_m = p_{ij}, \qquad \forall i, j \in A \tag{3}$$

3 Consistency conditions

We claim that the existence of a solution to the system of equalities (3) can be checked easily by verifying a condition on the p_{ij} values. We will prove both the sufficiency and necessity of this condition, and then finish this section by showing that the condition may be tested in polynomial time.

Theorem 1. *A dataset P can be rationalized by a mixture model of single-peaked preferences with respect to a given ordering $\gg$ if and only if for every triple $i, j, k \in A$ we have:*

$$\textit{if } i \gg j \gg k \textit{ then } p_{ij} \leq p_{ik} \textit{ and } p_{kj} \leq p_{ki} \tag{4}$$

It is easy to see that Condition (4) is a reformulation of Conditions (1-2) for the setting with aggregated preferences. While we will formally argue the necessity later on, it is clear that if condition (4) is violated, at least part of the population has to hold preferences that violate either condition (1) or (2). However, the sufficiency of this condition is not so straightforward. Indeed, if we look at mixture models with general preferences, we find that the list of necessary and sufficient conditions is exponential in the number of alternatives[1]. This is the case, even though general preference orderings are constrained only by transitivity, which can also be defined by a condition over all triples. To prove the sufficiency of condition (4) we will proceed as follows. First, we will describe an algorithm whose goal it is to find single-peaked preference orders $m \in O^{\gg}$ and associated values x_m satisfying (3). We will show this algorithm is able to do so if the dataset P satisfies condition (4).

The complete pseudo-code is given in Algorithm 1, here we will give a short overview. The main idea is that if we know the aggregated preferences of (a part of) a population, we wish to identify a single-peaked preference order held by a part of that population, explaining a portion

[1] Suck shows that testing the mixture model for strict linear preferences is equivalent to testing membership of the linear ordering polytope [10]. A full facet description of this polytope thus gives necessary and sufficient conditions for the mixture model of strict linear preferences. However, since separation over this polytope is NP-Hard, the facet description is exponential in the number of alternatives. Full descriptions are only known for small number of alternatives [6, 8]

of the observed preferences. This usually leaves some part of the data unexplained, captured in the algorithm by the variables $\tilde{p}_{ij}$. For this unexplained data, another single-peaked preference is then found, and so on. This process is the key part of the algorithm and is found in the loop (4-10). For this loop we will prove four properties, which all depend on the condition (4) being satisfied for $\tilde{p}_{ij}$. First, that the loop always run to completion, i.e. it outputs a strict linear order $\succ_m$. Second, that $\succ_m$ is single-single peaked with regards to $\gg$. Third, that this $\succ_m$ can be given a weight x_m, and for all $i, j \in A$ for which $i \succ_m j$, we have $x_m \leq \tilde{p}_{ij}$ and that there exist some $i, j \in A$ for which $i \succ_m j$ and $x_m = \tilde{p}_{ij}$. Finally, at the end of each loop, the values of $\tilde{p}_{ij}$ satisfy condition (4). Given these four properties, we will be able to prove that the algorithm provides single-peaked linear orders $m \in O^{\gg}$ and values x_m that satisfy (3).

Algorithm 1 Finding Single-Peaked Preferences

1: INPUT: p_{ij} for all $i, j \in A$ and $\gg$.
2: Set $\tilde{p}_{ij} := p_{ij}$ for all distinct $i, j \in A$, $m := 1$ and create $\succ_m := \emptyset$, $M := \emptyset$ and $I := \emptyset$.
3: **while** $\tilde{p}_{ij} + \tilde{p}_{ji} > 0$ for all distinct $i, j \in A$ **do**
4:   **for** $|M| < |A|$ **do**
5:     Set $I := \{i \in A \backslash M : \tilde{p}_{ij} \neq 0, \forall j \in A \backslash M, j \neq i\}$
6:     If $I = \emptyset$, STOP.
7:     Set $i^* := i$ with $i \in I$ for which $\forall j \in I, j \neq i : i \gg j$.
8:     $\forall j \in M$, set $j \succ_m i^*$
9:     Set $M := M \cup \{i^*\}$
10:   **end for**
11:   Set $x_m := \min_{i,j \in A: i \succ_m j} \tilde{p}_{ij}$.
12:   Set $\tilde{p}_{ij} := \tilde{p}_{ij} - x_m, \forall i, j \in A$ for which $i \succ_m j$.
13:   Set $m := m + 1$
14:   Set $M := \emptyset$
15: **end while**
16: OUTPUT: For all $i \in \{1, \ldots, m\}$ a value x_i and order $\succ_i$.

Claim 1. *If the values $\tilde{p}_{ij}$ meet condition (4), the loop (4-10) will return a linear order.*

Proof. If for some M there does not exist an $i \in A \backslash M$ such that $\tilde{p}_{ij} > 0, \forall j \in A \backslash M$, the algorithm will halt in line 6 without constructing an order. We argue by contradiction : suppose this is the case and the condition (4) is satisfied. Now consider $i \in A \backslash M$ with $i \gg j$ for all $j \in A \backslash M$. There is some j for which $i \gg j$ and $\tilde{p}_{ij} = 0$. Now let i' be the immediate neighbour of i [2]. Then by condition (4), we have $\tilde{p}_{ii'} = 0$. As $\tilde{p}_{ii'} + \tilde{p}_{i'i} > 0$ and thus, $\tilde{p}_{i'i} > 0$, (4) further implies $\tilde{p}_{i'l} > 0, \forall l \in A \backslash M$ for some $l \gg i'$. Furthermore, for i', there also exists some $j \in A \backslash M$ for which $\tilde{p}_{i'j} = 0$, this j must have $i' \gg j$. By the same argument as for i, we can see that $\tilde{p}_{i'i''} = 0$ and so on until we reach the alternative n, for which $j \gg n, \forall j \in A \backslash M$. This alternative n has $p_{nj} > 0, \forall j \in A \backslash M$, a contradiction. Given that condition 4 holds, there must exist an alternative which can be added to M in each step of the for loop, and the algorithm finds a strict linear order.

□

Claim 2. *If the values $\tilde{p}_{ij}$ meet condition (4), the linear order returned by the loop (4-10) is single-peaked with respect to $\gg$.*

[2]Meaning that there is no $k \in A \backslash M$ such that $i \gg k \gg i'$

Proof. First, we note that the set I has the following property. For each pair of alternatives $i, j \in I$, there does not exists a $k \notin I$ and $k \in A\backslash M$, for which $i \gg k \gg j$. This can be argued by contradiction, suppose such a k exists, then there also exists an alternative $l \in A\backslash M$ for which $\tilde{p}_{kl} = 0$. Without loss of generality we assume $l \gg k$. By conditions (4), $\tilde{p}_{lj} \geq \tilde{p}_{lk}$ and thus also $\tilde{p}_{jl} = 0$ in which case $j \notin I$.

Furthermore, consider an alternative j. In a given iteration of the loop, we have $j \in A\backslash M$, $j \notin I$, and $j \gg i$ for all $i \in I$. Only if there does not exist an alternative $j' \in A\backslash M$, $j' \notin I$, and $j \gg j' \gg i$, is $j \in I$ possible in the next iteration. Again we argue by contradiction, if $j \notin I$ in one iteration and $j \in I$ in the next, an alternative $i \in I$ with $\tilde{p}_{ji} = 0$ was added to M. If j' exists, condition (4) implies $0 = \tilde{p}_{ji} > \tilde{p}_{jj'}$ and j can not be added to I. The same argument applies for $i \gg j$

From these two observations the claim easily follows. Suppose $i \in A$ is the first alternative set for order m. For every pair of alternatives $j, k \in A$, for which $j \gg k \gg i$, k is added to the order m before j, as the iteration in which $j \in I$ must be after the iteration in which $k \in I$ and k will be added to the order immediately when $k \in I$. For every pair of alternatives $h, l \in A$, for which $i \gg h \gg l$, h is added to the order before l as $l \in I$ implies $h \in I$ and if both $l, h \in I$, l can not be chosen as $h \gg l$ by construction in line 7 $\square$

Claim 3. *In line 11, x_m is set such that $x_m \leq \tilde{p}_{ij}$ for all $i, j \in A$ for which $m \in O^{\gg}_{ij}$ and there exists some $i, j \in A$ for which $m \in O^{\gg}_{ij}$, such that $x_m = \tilde{p}_{ij}$ and $x_m > 0$.*

Proof. This is true by construction, an alternative i is only added to M if $\forall j \in A\backslash M, \tilde{p}_{ij} > 0$. As $i \succ_m j$ is only the case if j was added to M after i, then all $\tilde{p}_{ij}$ over which the minimization are done are strictly positive. By nature of the minimization, there is also at least one $\tilde{p}_{ij}$ to which x_m is equal and x_m is no larger than any of the $\tilde{p}_{ij}$. $\square$

Claim 4. *If condition (4) is satisfied at the beginning of the loop (3-15), the $\tilde{p}_{ij}$ values will satisfy condition (4) at the end of the loop.*

Proof. In this proof, we will denote the value $\tilde{p}_{ij} + \tilde{p}_{ji}$ by y. Throughout, we will assume that condition (4) is satisfied in line 3. First, let us consider the situation $i \gg j \gg k$, as (4) holds, $\tilde{p}_{ij} \leq \tilde{p}_{ik}$. Only if an order m exists such that $j \succ_m i \succ_m k$ is found, will $\tilde{p}_{ik}$, but not $\tilde{p}_{ij}$, decrease in line 11. If both $i, j \in I$, i will be added to m first due to line 7. Thus, $j \succ_m i$ implies that there exists $l \in A$, such that $\tilde{p}_{il} = 0$ and $\tilde{p}_{jl} > 0$. We will consider three distinct situations. First, $l \gg i$, then $j \gg l$ and finally $i \gg l \gg j$. Let us consider $l \gg i$. As $\tilde{p}_{il} = 0$, $\tilde{p}_{li} = y$, which implies $\tilde{p}_{lj} = y$ and $\tilde{p}_{jl} = 0$ as $l \gg i \gg j$ demands $\tilde{p}_{li} \leq \tilde{p}_{lj}$. Therefore, l would prevent both i and j from being added to I and m. As soon as l was added to M, both $i \in I$ and $j \in I$ are possible, and again i would be added to m before j. $l \gg i$ thus can not lead to $j \succ_m i$. In the case of $j \gg l$, it is clear that because $\tilde{p}_{il} = 0$, we must also have $\tilde{p}_{ij} = 0$, therefore $\tilde{p}_{ij} \leq \tilde{p}_{ik}$ can not be violated. Finally, if $i \gg l \gg j$, we must have $j \succ_m l$, if this were not the case i could be added to m after l but before j. By the earlier arguments in this paragraph $j \succ_m l$ while $l \gg j$ is only possible if there is some other alternative $l' \in A$, with $l \gg l'$ and $\tilde{p}_{ll'} = 0$. $j \gg l'$ gives $\tilde{p}_{lj} = 0$ and therefore $\tilde{p}_{ij} = 0$. If on the other hand $l \gg l' \gg j$, we can repeat the same argument until we find some l'' with $i \gg l'' \gg j$ and $\tilde{p}_{jl''} = y$, implying $\tilde{p}_{ji} = y$ and $\tilde{p}_{ij} = 0$. In conclusion, if $i \gg j \gg k$, we can only have $j \succ_m i \succ_m k$ if $\tilde{p}_{ij} = 0$. If this is the case, then $\tilde{p}_{ij} \leq \tilde{p}_{ik}$ is satisfied, as $\tilde{p}_{ik} \geq 0$.

The second situation is $k \gg j \gg i$, in which case we also have $\tilde{p}_{ij} \leq \tilde{p}_{ik}$. Here, only an order with $j \succ_m i \succ_m k$ can lead to the condition being violated in line 15. In the previous paragraph, we established that if $a \gg b$ and the algorithm places $b \succ_m a$, we have $\tilde{p}_{ab} = 0$. Here, $k \gg i$ and

$i \succ_m k$, so $\tilde{p}_{ki} = 0$. As $\tilde{p}_{ki} = 0$, it must be the case that $\tilde{p}_{ik} = y$ and thus $\tilde{p}_{ij} \leq y = \tilde{p}_{ik} = y$. □

We are now in a position to prove theorem 1.

Proof. First, we prove sufficiency of the condition. We have shown, by combining claims 1 and 2, that given a set of values $\tilde{p}_{ij}$ which satisfy condition (4), we can find a strict single-peaked linear order. By claim 3 we have also seen that we can attach a weight to this order which is non-negative. Even stronger, we have shown that this weight is equal or less than the value $\tilde{p}_{ij}$ for some $i, j \in N$, for which $x_m \in O_{ij}^{\gg}$. As the final step of the loop will decrease these $\tilde{p}_{ij}$ values, at least one of these values is set to zero in each run. After at most $O(n^2)$ iterations of the loop, each value $\tilde{p}_{ij}$ will then be zero. It can be easily checked that at this point, the values x_m form a solution to (3). As this proof requires the loop to be run multiple times, and the loop requires condition (4) to hold, claim 4 is crucial, as it shows that if the input of the loop satisfies the condition, the output will as well.

Next, we turn to the necessity of condition (4). This can easily be verified by a three alternative example. Suppose $i, j, k \in A$, with $i \gg j \gg k$ and $p_{ij} > p_{ik}$. By definition of single-peaked linear orders, each order for which $i \succ j$ also has $i \succ k$. This means $O_{ij}^{\gg} \subset O_{ik}^{\gg}$ and $\sum_{m \in O_{ij}^{\gg}} x_m \leq \sum_{m \in O_{ik}^{\gg}} x_m$. A solution to (3) requires $p_{ij} = \sum_{m \in O_{ij}^{\gg}} x_m$ and $p_{ik} = \sum_{m \in O_{ik}^{\gg}} x_m$, but this is obviously impossible. The same argument can be used for $p_{kj} \leq p_{ki}$. This shows necessity of the condition. □

Theorem 2. *For a given dataset P and ordering $\gg$, Condition (4) may be checked in time $O(n^2)$.*

Proof. It can be easily seen that this condition may be checked in polynomial time. As written, two inequalities must be checked for each triplet of alternatives, giving an obvious $O(n^3)$ time test. This can be improved upon by noting that when using a matrix of p_{ij} values, with rows and columns ranked according to the ordering $\gg$, values above the diagonal must be non-decreasing in the rows and the columns. Conversely, as $p_{ij} + p_{ji} = 1$, values below the diagonal are non-increasing in both rows and diagonals[3]. As such, each p_{ij} value must be compared with only two other values, providing an $O(n^2)$ test. □

4 Identifying the single-peaked ordering

In the previous section, we have given necessary and sufficient conditions for the data to be consistent with a mixture model of single-peaked preferences with respect to an order $\gg$. In this section, we will show that if such an order is not given a priori, but there do exist orderings for which the dataset P satisfies the mixture model, we can identify these. We define the set of orders L_P, with $\gg \in L_P$ if and only if P satisfies condition (4) with respect to $\gg$. In this section, we will prove that we can identify L_P. Note that if an order $\gg \in L_P$, the reverse order $\ll \in L_P$. This can be easily checked because the condition (4) depends on the relative ordering of alternatives, but not its orientation. $i \gg j \gg k$ and $i \ll j \ll k$ lead to the same constraints on P.

Initially, we make the assumption that there is no subset $A' \subset A$ with $|A'| > 1$, for which the following holds: $\forall i, j \in A', k \in A \backslash A', p_{ik} = p_{jk}$. In words, this means that there is no subset of A with two or more items for which all items seem identical when compared to items outside of

[3] In fact, the lower triangle of the matrix is double graded

this subset. We will call these *Nearly-Identical Alternative Subsets* (NIA Subsets) and show how to recognize and handle such subsets in subsection 4.1.

We start with the special case without any NIA subsets and proceed as follows. First, we derive a number of necessary conditions for all $\gg \in L_P$. If satisfied, the second and third of these conditions, given in Claim 6 and 7, can be used to identify an *extreme alternative*, $\bar{a}$, which is either the first or last element of any ordering $\gg \in L_P$. This extreme alternative is then used as input for Algorithm 2 and we will show that, if $L_P \neq \emptyset$, this algorithm has an ordering $\gg \in L_P$ as output. We begin by deriving another necessary condition on the p_{ij} values, which we will use in further proofs.

Claim 5. *For any $\gg \in L_P$ and each triple of alternatives $i, j, k \in A$:*

$$\text{if } i \gg j \gg k \text{ then } p_{ij} \leq p_{jk} \tag{5}$$

Proof. Suppose this is not the case and $p_{jk} < p_{ij}$. Due to condition (4), we further have $p_{jk} < p_{ij} \leq p_{ik}$ and $p_{kj} \leq p_{ki}$. Equivalently, $1 - p_{jk} \leq 1 - p_{ik}$ or $p_{jk} \geq p_{ik}$, which violates condition (4). □

The next claim gives a further necessary condition for L_P to be non-empty. Specifically, L_P can only be non-empty if all alternatives are indistinguishable from each other, or if there exists an extreme alternative $\bar{a}$ which is a Condorcet-loser.

Claim 6. *Suppose there exists an ordering $\gg \in L_P$. Then either*

$$p_{ij} = 0.5, \forall i, j \in A \tag{6}$$

Or

$$\exists \bar{a} : \{p_{\bar{a}i} \leq 0.5, \forall i \in A \text{ and } \exists j \in A : p_{\bar{a}j} < 0.5\} \tag{7}$$

Proof. First, suppose there is no extreme alternative $\bar{a}$ for which for all $i \in A$ $p_{\bar{a}i} \leq 0.5$. Without loss of generality, we say $p_{a_1 i} > 0.5$. Then, because $a_1 \gg i \gg a_n$ and condition (4), we have $p_{a_1 a_n} > 0.5$, thus $p_{a_n a_1} < 0.5$ and again by condition (4), we have for all $j \in A$, $p_{a_n j} < 0.5$. Thus, there is certainly at least one extreme alternative for which for all $i \in A$, $p_{\bar{a}i} \leq 0.5$.

Now suppose an extreme alternative $\bar{a}$ exists for which for all $i \in A$, $p_{\bar{a}i} \leq 0.5$, but there exists no extreme alternative for which there exists an $i \in A$ for which $p_{\bar{a}i} < 0.5$. Without loss of generality we assume that this is the case for a_1 , thus for all $i \in A$, $p_{a_1 i} = 0.5$. This includes $p_{a_1 a_n} = 0.5$, so by condition (4), we also have $p_{a_n i} \leq 0.5$ for all $i \in A$. As we assume there is no extreme alternative for which there exists an $i \in A$ for which $p_{\bar{a}i} < 0.5$, we also have $p_{a_n i} = 0.5$ for all $i \in A$. Now consider $i, j \in A$, with $a_1 \gg i \gg j \gg a_n$. Then condition (5) and $p_{a_1 i} = 0.5$ imply $p_{ij} \geq 0.5$ and the same condition and $p_{a_n j} = 0.5$ imply $p_{ji} \geq 0.5$. As a result, for each pair $i, j \in A$, we have $p_{ij} = 0.5$. □

Condition (7) shows a way to identify an extreme alternative $\bar{a}$ for any ordering $\gg \in L_P$. However, it remains to be shown that no non-extreme alternatives share this characteristic, which we do in the next Claim.

Claim 7. *For any ordering $\gg \in L_P$, there is no non-extreme alternative a for which for all $i \in A$, $p_{ai} \leq 0.5$ holds and there exists $j \in A$ for which $p_{aj} < 0.5$.*

Proof. Suppose such an alternative exists. Without loss of generality, suppose a_1 is an extreme alternative satisfying the condition in Claim 6 and $a_1 \gg a \gg a_n$. Then $p_{a_1 a} = 0.5$ and $p_{a_n a} \geq 0.5$. By condition 5, we then have for all $i \in A$ for which $a_1 \gg a \gg i$, $p_{ai} \geq p_{a_1 a} = 0.5$ and for all $j \in A$ for which $j \gg a \gg a_n$, $p_{ai} \geq p_{a_n a} \geq 0.5$. Thus, such an alternative cannot exist. □

We are now in a position to describe our algorithm for identifying the orders $\gg \in L_P$. As an initial step, we will check the conditions described in Claim 6. If neither (6) or (7), $L_P = \emptyset$. Furthermore, it is easy to see that if condition (6), P satisfies the mixture model with respect to any ordering $\gg$. In other words, L_P is the set of all linear orders over the alternatives. Finally, if an alternative a matching condition (7) is found, Claims 6 and 7 prove it is an extreme alternative $\bar{a}$ for any $\gg \in L_P$. With $\bar{a}$ as input, Algorithm 2 can now be used to identify a complete ordering $\gg \in L_P$, provided such an ordering exists.

The main idea of Algorithm 2 is as follows. Given an extreme alternative $\bar{a}$ and two alternatives $i, j \in A$ for which $p_{\bar{a}i} \neq p_{\bar{a}j}$, the relative ordering of i and j is determined, such that Condition (4) is satisfied for the triple $\bar{a}, i, j$. When this is done for every pair of alternatives, this results in a (partial) order of the alternatives. For any pair of alternatives, for which $p_{\bar{a}i} = p_{\bar{a}j}$, a third alternatives $k \in A$ is sought, which in the partial order has $k \gg i, j$ or $k \ll i, j$ and for which $p_{ki} \neq p_{kj}$. In this way, the partial order is refined until a full ordering of the alternatives is found.

Algorithm 2 Ordering Algorithm

1: Input: Dataset P, set A, extreme alternative a_1
2: Create ordering $\gg$
3: For every $i \in A \backslash \{a_1\}$, set $a_1 \gg i$
4: For each pair $i, j \in A$, for which $p_{a_1 j} > p_{a_1 i}$, set $i \gg j$.
5: Divide all $i \in A$ into sets $A^1, A^2, \ldots$, such that for all $i, j \in A^k$ neither $i \gg j$ or $i \ll j$
6: **repeat**
7: Find a set A^k with $\left|A^k\right| > 1$, for which there exist $i \in A \backslash A^k$ and $j, j' \in A^k$ such that $p_{ij} > p_{ij'}$
8: **if** $i \gg j$ **then**
9: For each pair $j, j' \in A^k$ for which $p_{ij} > p_{ij'}$, set $j' \gg j$
10: **else**
11: For each pair $j, j' \in A^k$ for which $p_{ij} > p_{ij'}$, set $j \gg j'$
12: **end if**
13: **if** There exists a pair $j, j' \in A^k, j \gg j'$ and $i \in A \backslash A^k$, for which Condition (4) is violated **then**
14: Algorithm terminates without OUTPUT.
15: **end if**
16: Divide all $i \in A$ into sets $A^1, A^2, \ldots$, such that for all $i, j \in A^k$ neither $i \gg j$ or $i \ll j$
17: **until** For all $i = 1, \ldots, n$, $\left|A^i\right| = 1$
18: OUTPUT: An ordering of the alternatives $\gg$.

Claim 8. *If there are no Nearly-Identical Alternative Subsets, Algorithm 2 terminates.*

Proof. First, note that if there exists a set A^k with $\left|A^k\right| > 1$, for which there exist $i \in A \backslash A^k$ and $j, j' \in A^k$ such that $p_{ij} > p_{ij'}$, a (partial) ordering is made of the alternatives in this set and A^k is thus further split up in line (16). As a result, if such a set is always found at the start of the loop (6-17), the stopping condition for this loop (line 17) is reached in a finite number of iterations. Now suppose Algorithm 2 does not terminate, then there is a set A^k with $\left|A^k\right| > 1$, and there exists no $i \in A \backslash A^k$, such that there exist $j, j' \in A^k$ for which $p_{ij} \neq p_{ij'}$. By definition, this set is a nearly-identical alternative subset. □

Claim 9. *If there are no Nearly-Identical Alternative Subsets, an ordering $\gg \in L_P$ exists and the alternative a_1 is an extreme alternative of this ordering, then Algorithm 2 terminates with output $\gg$ and $\gg \in L_P$.*

Proof. Given that an ordering $\gg \in L_P$ exists and a_1 is an extreme alternative, it must be the case that $a_1 \gg i$ for all $i \in A$. Furthermore, it can be easily checked than whenever the relative ordering of two alternatives $i, j \in A$ is fixed in relation to a third alternative $k \in A$ (say $k \gg i \gg j$), whether in line (4) or the loop (6-17), the opposite relative ordering $k \gg j \gg i$ would violate Condition (4). □

Claim 10. *If there are no Nearly-Identical Alternative Subsets and there exists an ordering $\gg \in L_P$ $|L_P| = 2$*

Proof. This follows immediately from the previous result. If no NIA Subsets exist, Algorithm 2 terminates with output $\gg inL_P$, and whenever a relative ordering of two alternatives is fixed, the opposite ordering would violate Condition (4). However, if $\gg \in L_P$, it can be easily checked that the reverse order $\ll \in L_P$, thus $L_P = \{\gg, \ll\}$. □

4.1 Nearly-Identical Alternative Subsets

We have now shown how to identify the ordering $\gg$, for which the data satisfy a mixture model of single-peaked preferences, under the assumption that there are no NIA subsets. In this subsection, we will show how to handle such subsets.

As a starting point, we will again look at Algorithm 2. If NIA subsets are present, at some point it will be impossible to find a subset $\left|A^k\right| > 1$ in line (7) to split up in the loop (6-17). Without loss of generality, let us assume there is a single subset $|A'| > 1$, then there is a partial ordering of the alternatives $a_1 \gg \ldots \gg a^- \gg A' \gg a^+ \gg \ldots \gg a_n$. It can be easily proven, by a similar argument as for Claim 9, that for any triple $i, j, k \in A \backslash A'$ and any triple $i, j \in A \backslash A'$, $k \in A'$, Condition (4) is satisfied. What remains to be shown is that we can extend $\gg$, such that Condition (4)is satisfied for any triple $i \in A \backslash A'$, $j, k \in A'$ and any triple $i, j, k \in A'$.

Consider all pairs $i, j \in A'$, such that $\max_{r,s \in A'}(p_{rs}) = p_{ij}$. If $p_{ij} > p_{ia^-}$ and $p_{ij} > p_{ia^+}$, it is clear that no $\gg \in L_P$ exists, as both $a^- \gg j \gg i \gg a^+$ and $a^- \gg i \gg j \gg a^+$ violate Condition (4). Next, if (without loss of generality) $p_{ij} > p_{ia^-}$, but $p_{ij} \leq p_{ia^+}$, it must be the case that $a^- \gg i \gg j$. Furthermore, there is no $k \in A'$, $\max_{r,s \in A'}(p_{rs}) \neq p_{ik}$ such that $a^- \gg i \gg j \gg k$, as this would also violate Condition (4). As a result, for all $j \in A'$, for which there exists $i \in A'$, $\max_{r,s \in A'}(p_{rs}) = p_{ij}$ and all $k \in A'$ for which there does not exist $i \in A'$, $\max_{r,s \in A'}(p_{rs}) = p_{ik}$, it must be the case that $k \gg j$. At this point, the subset A' is split into two subsets, and Algorithm 2 can be resumed. Finally, if both $p_{ij} \leq p_{ia^-}$ and $p_{ij} \leq p_{ia^+}$, there can be no violation of Condition (4) for any triple $r \in A \backslash A'$, $s, t \in A'$. Now, the question is whether there exists an ordering $\gg' \in L_{P'}$. This question can again be answered using Algorithm (4). If such an order is found, then both $\gg'$ and it's reverse $\ll'$ can be used to complete the partial order $\gg$, in other words, the ordering $a_1 \gg \ldots \gg a'_1 \gg' \ldots \gg' a'_n \gg \ldots \gg a_n$ and $a_1 \gg \ldots \gg a'_n \gg' \ldots \gg' a'_1 \gg \ldots \gg a_n$ are in L_P. In this case, we will call the NIA subset a *re-orientable* subset. Denote the total number of re-orientable subsets by R, then we finish with the following result.

Claim 11. *If $L_P \neq \emptyset$, then $|L_P| = 2^{R+1}$.*

Proof. We have shown that for every re-orientable subset and a given partial order over alternatives outside of this subset, there are two ways to complete that partial order that satisfy Condition (4). Thus, every re-orientable subset doubles the number of complete orders $\gg \in L_P$. $\square$

5 Conclusion

In this paper, we present a mixture model from the choice behaviour literature and apply it to a well-known choice domain from the social choice literature. Necessary and sufficient conditions are derived for the mixture model to hold for single-peaked preferences and a given ordering of the alternatives. Furthermore, we show that these conditions are easy to check in polynomial time, in contrast to the mixture model for general preferences. Furthermore, a polynomial time algorithm is provided to identify whether or not there exists some ordering of the alternatives for which the mixture model is satisfied.

References

[1] M. Ballester and G. Haeringer. A characterization of the single-peaked domain. *Social Choice and Welfare*, 36(2):305–322, 2011.

[2] J. Bartholdi III and M. Trick. Stable matching with preferences derived from a psychological model. *Operations Research Letters*, 5(4):165–169, 1986.

[3] D. Black. On the rationale of group decision-making. *The Journal of Political Economy*, 56(1):23, 1948.

[4] C. Davis-Stober. A lexicographic semiorder polytope and probabilistic representations of choice. *Journal of Mathematical Psychology*, 56(2):86–94, 2012.

[5] B. Escoffier, J. Lang, and M. Öztürk. Single-peaked consistency and its complexity. In *ECAI*, volume 8, pages 366–370, 2008.

[6] S. Fiorini. 0, 1/2-cuts and the linear ordering problem: Surfaces that define facets. *SIAM journal on discrete mathematics*, 20(4):893–912, 2006.

[7] V. Knoblauch. Recognizing one-dimensional euclidean preference profiles. *Journal of Mathematical Economics*, 46(1):1–5, 2010.

[8] R. Martí and G. Reinelt. *The Linear Ordering Problem: Exact and Heuristic Methods in Combinatorial Optimization*, volume 175 of *Applied Mathematical Sciences*. Springer-Verlag Berlin Heidelberg, 2011.

[9] M. Regenwetter, J. Dana, and C.P. Davis-Stober. Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in psychology*, 1(148):1–15, 2010. doi: 10.3389/fpsyg.2010.00148.

[10] R. Suck. Geometric and combinatorial properties of the polytope of binary choice probabilities. *Mathematical Social Sciences*, 23(1):81–102, 1992.

[11] M. Trick. Recognizing single-peaked preferences on a tree. *Mathematical Social Sciences*, 17(3):329–334, 1989.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be